THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 38 February, 1947 Number 2

MENTAL GROWTH AND DEVELOPMENT AT THE COLLEGE LEVEL

SISTER M. FLORENCE LOUISE
Marygrove College, Detroit

As a part of a program designed to evaluate objectively the intellectual growth and development of its students, Marygrove College, since 1939, has administered to the senior class the current edition of the American Council on Education Psychological Examination. The same test, in an equivalent form, is taken by each student during orientation week at the beginning of the freshman year. Since there is an interval of three and a half years, or to be exact of forty months, between the administration of the first and the second tests and since the forms used are not identical, the practice effect, presumably, is negligible. The same examiner has administered both tests for the past ten years, so it is extremely probable that the conditions of administration were about identical for each group reported in this study. The retests are given to each class at the end of the first semester of the senior year, during the week of the first semester examinations. The period intervening between the test and the retest, therefore, allows for the influence of seven semesters of college experience.

Marygrove is a liberal arts college for women located in Detroit, Michigan, enrolling both resident and non-resident students. Because of the lack of residential facilities, the non-resident enrollment is far in advance of the resident. Thus, during the scholastic year 1945–1946 only one hundred ninety-seven of the seven hundred students registered lived on campus; the others were residents of Detroit and the surrounding towns and commuted back and forth to college daily.

The present study shows the growth of the students of the

classes of 1943, 1945, and 1946 as measured by improvement in the scores obtained on the A. C. E. tests after an interval of seven semesters. Results were not available for the class of 1944 since it was impossible to administer the retest to this group at the usual time. As freshmen the class of 1943 were given the 1939 test and as seniors the 1942 edition; the class of 1945 took the 1941 edition as freshmen and the 1944 test as seniors; the 1942 edition was given to the class of 1946 when they were freshmen and the 1945 edition in the middle of the senior year. Only students who took both tests at the times assigned for the group are included in the study. For this reason the records of seniors who entered their respective classes late in the freshman year or as juniors or seniors or who had spent five years in acquiring a college degree were eliminated from the investigation. Percentile ranks were computed from the norms published by the American Council on Education each year. Raw scores for the various editions of the test are not directly comparable, especially between the 1939 edition, which apparently is more difficult, and the others, but both raw scores and percentile ranks of the different editions may be compared directly through the data published by the Council annually. In this study, in every case, norms are based on the scores derived from the machine-scored edition of the test, the only edition now published.

TABLE I.—RESULTS OF FRESHMEN TESTS (TOTAL SCORE)

Class	Number of Cases	Test Used	Mean Score	$SD_{dist.}$	SD_m	Percentile Rank
1943	107	1939 edition	81.70 (101)	16.76	1.62	47
1945	86	1941 edition	104.24	22.71	2.45	48
1946	95	1942 edition	104.47	24.25	2.50	51

There have been a few previous studies somewhat related to the one here reported. No attempt will be made to review these investigations but comparisons between the results of these and the present study will be given. Table I shows the mean, standard deviation, the standard error of the mean, and percentile rank for the total score of each class in the freshman year. Table II gives the same data for the retests given during the senior-year. Table III summarizes the gain in raw score, gain in percentile points, the standard error of the difference of

the means, and the critical ratio for each of the three classes included in the study.

TABLE II.—RESULTS OF SENIOR TESTS (TOTAL SCORE)

Class	Number of Cases	Test Used	Mean Score	$SD_{dist.}$	SD_m	Percentile Rank
1943	107	1942 edition	119.34	18.15	1.76	75
1945	86	1944 edition	123.49	21.30	2.30	75
1946	95	1945 edition	120.68	22.58	2.32	77

TABLE III.—GAIN IN RAW SCORE AND PERCENTILE POINTS, $\sigma_{diff.\ of\ means}$, CRITICAL RATIO (TOTAL SCORE)

Class	Number of Cases	Gain in Raw Score	Gain in Percentile Points	σ Diff.$_m$	Critical Ratio
1943	107	18.34	28	2.39	7.67
1945	86	19.25	27	3.36	5.73
1946	95	16.21	26	3.41	4.75

Tables I, II, and III show definitely that each of the three groups made considerable improvement in the total score from the freshman to the senior year. The raw scores for the 1939 test are low, due, as stated before, to the difference in the difficulty of that edition, so the norms for this test were computed from data published by the Council showing that a score of 81 on the 1939 edition of the test is equivalent to 101 on the 1942 edition. In computing the difference in the raw scores for the two tests of the Class of 1943, therefore, 101, instead of 81.70, was used as the mean score for the freshman year. Table I shows that the students included in this study were fairly typical of freshmen throughout the nation as they ranked in the 47th, 48th and 51st percentiles on the gross score. The fact that these percentiles were raised to 75, 75, and 77 in the senior year, with critical ratios larger than four in each case, proves that there were statistically significant gains made by each of the three classes during the three-and-a-half-year interval between test and retest. Percentile point gains of 28, 27, and 26 are similar to those found by Hunter[5] in his study, reported in 1942, of forty-three women students at Converse College, South Carolina. He found a percentile gain from freshman to senior year of thirty-one

points for students ranking in the 48, 49, 51, and 62 percentiles as freshmen. In the present study the standard deviations for the senior year, in every case but one, are less than for the freshman year, but the differences are so slight that it must be concluded that the older students, as a group, are almost as heterogeneous as the younger. The problem of individual differences, therefore, appears to be about as great in the upper as in the lower division of the college. These findings agree with those of Hunter[5] and Flory.[2] The former concluded from the slight standard variation in scores on the readministration of the A. C. E. tests that "college experience, regardless of amount, does little to reduce the wide variability in test scores of students."

Since 1938 the A. C. E. tests have been arranged, and the data tabulated, so that it is possible to obtain two significant subscores, the Q-score derived from the three tests of a quantitative nature and the L-score obtained from the three tests that are largely linguistic in content. The two subscores represent what the Council calls, "two groups of abilities that are significant for college curricula that are dominantly language or technical."

TABLE IV.—RESULTS OF FRESHMAN TESTS (Q-SCORE)

Class	Number of Cases	Test Used	Mean Score	$SD_{dist.}$	SD_m	Percentile Rank
1943	107	1939 edition	29.46 (36)	8.90	.86	38
1945	86	1941 edition	39.70	9.11	.98	45
1946	95	1942 edition	37.02	9.05	.93	39

TABLE V.—RESULTS OF SENIOR TEST (Q-SCORE)

Class	Number of Cases	Test Used	Mean Score	$SD_{dist.}$	SD_m	Percentile Rank
1943	107	1942 edition	40.91	8.40	.81	54
1945	86	1944 edition	44.15	9.09	.98	70
1946	95	1945 edition	41.60	8.70	1.00	60

They are valuable in counseling students concerning curricula in which success is probable. Tables IV, V, and VI summarize results for the Q-scores equivalent to those given in Tables I, II, and III for the total score.

The data in Tables IV, V, and VI, and those in the three tables that follow, show again that the students included in the study are typical of women students in general, since they rank lower in quantitative than in linguistic scores. Though they fall below the 50th percentile in Q-score for the freshman year, still without exception each group is well above the average percentile in the senior year. The Class of 1945 gathered the

TABLE VI.—GAIN IN RAW SCORE, GAIN IN PERCENTILE POINTS FROM FRESHMAN TO SENIOR YEAR, $SD_{diff.m}$, AND CR (Q-SCORE)

Class	Number of Cases	Gain in Raw Score	Gain in Percentile Points	$SD_{diff.m}$	CR
1943	107	4.91	16	1.18	4.16
1945	86	4.45	25	1.39	3.20
1946	95	4.58	21	1.36	3.37

greatest gain in percentile points on Q-score finally reaching the 70th percentile, thereby averaging a gain of twenty-five percentile points with a critical ratio of 3.20 proving that the difference is significant. It is interesting to note that this class reached the highest percentile as freshmen, as compared with the other two groups, and that it gained the most.

This is not true of the L-score, as Table IX shows. The class of 1946 reached the 62nd L-score percentile as freshmen, but gained only twenty-three percentile points as compared with gains of twenty-eight and thirty-three points by groups that ranked in the 56th and 50th percentiles on the L-score as freshmen. In the case of the L-score, therefore, the class that ranked the lowest gained the most, whereas in Q-score the class that ranked the highest increased the score the most. The large average gain in Q-score of the Class of 1945 may be due to the fact that only three students in that class lost a considerable number of points during the interval between test and retest as compared with seven and fourteen persons, respectively, in the other groups whose scores decreased between ten and thirty-nine points. Furthermore only twelve students of this group lost percentile points on the retest as compared with twenty-six and twenty-three, respectively, in the classes of 1943 and 1946. Then, too, the increase may be accounted for by the fact that the Class of 1945, a war class, had more students majoring in

mathematics and science than the usual Marygrove class. Though most of the mathematics and science majors made relatively large gains, still, on the other hand, Latin, Spanish, Art, Music, History, and English majors increased their Q-scores enough to rank in the top decile, while two Science majors actually lost in terms of percentile points. The large average gain made in Q-score by the three groups, therefore, must be assigned not only to specific training in quantitative thinking, but also to mental growth occurring during the college years, for most of the students had but one year of mathematics or science, taken usually in the freshman class and in many cases the science chosen was Biology. The gain is too large to ascribe to emotional or physical tensions affecting the score of the first test nor to incidental learning which probably influenced some retest scores.

The gain in Q-score found in the three groups studied here is larger than that found by Rogers[8] for fifty-six college women at Bryn Mawr. Rogers reports that students having no mathematical instruction in college lost on an average of 5.4 points on the readministration of the Thorndike test. Those majoring in mathematics gained 5.4 points, while those having some mathematical instruction, but not majoring in it, gained 1.8 points. The difference in the structure of the two tests may account partially for the difference in the findings, but the fact remains that more than 77.4 per cent of the two hundred eighty-eight students in this study gained on the average of four and a half points in raw score and that 1.4 per cent kept the same score; only 21.2 per cent of this group lost points in quantitative ability during the forty-month interval.

As in the case of the gross scores, the standard deviations for the Q-scores for the senior year vary only slightly from the SD of the same groups for the freshman year, showing that in quantitative ability the classes were only slightly more homogeneous as seniors than as freshmen.

Tables VII, VIII, and IX give similar data for the L-score for the three classes.

Table VII shows that the freshmen admitted to Marygrove are average or above in linguistic ability and that each year the groups appear to become more select, the mean ratings reaching the 50th, 56th, and the 62nd percentiles. These ratings increase

to the 83rd, 84th, and 85th percentiles in the senior year, indicating gains of thirty-three, twenty-eight, and twenty-three percentile points in L-scores during the seven semesters of college influence. The large critical ratios indicate that the differences are statistically significant. Standard deviations in two of the three classes are slightly higher in the senior than in the freshman year, again pointing to the need of proper guidance, appropriate motivation, and provision for individual differences in the

TABLE VII.—RESULTS OF FRESHMEN TEST (L-SCORES)

Class	Number of Cases	Test Used	Mean Score	$SD_{dist.}$	SD_m	Percentile Rank
1943	107	1939 edition	51.15 (62)	11.25	1.09	50
1945	86	1941 "	67.45	15.72	1.70	56
1946	95	1942 "	67.84	15.90	1.63	62

TABLE VIII.—RESULTS OF SENIOR TESTS (L-SCORES)

Class	Number of Cases	Test Used	Mean Score	$SD_{dist.}$	SD_m	Percentile Rank
1943	107	1942 edition	78.50	12.35	1.20	83
1945	86	1944 "	80.35	14.15	1.53	84
1946	95	1945 "	80.37	16.00	1.64	85

TABLE IX.—GAINS IN RAW SCORES AND PERCENTILE POINTS, $SD_{diff.}$ AND CRITICAL RATIOS (L-SCORES)

Class	Number of Cases	Gain in Raw Scores	Gain in Percentile Points	$SD_{diff.}$	Critical Ratio
1943	107	16.50	33	1.62	10.19
1945	86	12.90	28	2.29	5.63
1946	95	12.53	23	2.31	5.42

presentation of material in the upper college. The drop in gain in percentile points from year to year may be explained, partially at least, by the fact that the freshman group each year reached a higher percentile rank and it has been fairly well established by this present study, by Flory[2] and Hunter[5] that gains are less in the higher intellectual brackets than in the middle and lower. It stands to reason that in the freshmen top decile group there is little chance to improve, and for some students, who as fresh-

men, reach the 99th percentile, no opportunity. There is little doubt, however, that there is a gain that the tests do not measure. This fact suggests the need of a test that will measure and discriminate the various mental levels of the gifted college or adult student. In the field of higher education at the present time there is no group test that measures these students adequately.

To uncover gains and losses hidden by mean scores the data were further studied to determine the number of students who increased, lost, or kept the same percentile rank they had as freshmen. It was found that the vast majority of the students increased their total score during the college years. One hundred five (ninety-eight per cent), eighty-two (ninety-five per cent), and ninety (ninety-five per cent), or two hundred seventy-seven individuals of the two hundred eighty-eight in the three groups displayed augmented intellectual ability on the retest. No student in the Class of 1943 lost percentile points on the total score, but in the classes of 1945 and 1946 three students in each group lost—a total of only 2.1 per cent of the two hundred eighty-eight included in the study. After an interval of seven semesters only five students in the three groups kept the same total score, three because they had reached the 99th percentile as freshmen. One hundred five, eighty-two, and eighty-six, totaling two hundred seventy-three or 94.8 per cent of the two hundred eighty-eight students tested, gained in L-score points. Six students, four because they had reached the 99th percentile on the first test, kept the same L-score. The number of students losing percentile points on the L-score was small, totaling nine or three per cent for the three years. Losses on the Q-score were more numerous—twenty-six, twelve, and twenty-three, a total of sixty-one students or 21.2 per cent for the three classes investigated. On the other hand seventy-nine, seventy-three, and seventy-one, a total of two hundred twenty-three students or 77.4 per cent of the total number, gained in Q-score percentile points. Four students, about 1.4 per cent of the entire group, kept the same Q-score rating from test to retest.

To determine the amount and direction of change of ratings from the freshman to the senior year the data were further analyzed. Many increases on the gross score were considerable. A total of five, eight, and one (or fourteen) students increased the total rating by fifty or more percentile points; seventeen per cent,

fourteen per cent and nine per cent gained forty or more percentile points; thirty-eight per cent, twenty-eight per cent, and twenty-one per cent increased their rank by thirty or more percentile points; sixty-three per cent, fifty-three per cent, and fifty-two per cent gained twenty or more; eighty-eight per cent, seventy-six per cent, and seventy-two per cent gained ten or more; and ninety-eight per cent, ninety-five per cent, and ninety-five per cent gained one or more percentile points. The relatively large number of students (thirty-eight, thirty-six and forty-one) making the smaller gains of from one to nineteen points were chiefly those who scored in the top or second decile on the first test and for whom large gains were, therefore, impossible. In the Class of 1945 one student gained seventy-three and another seventy points. During the upper college years both of these girls displayed remarkable growth and development. Though the argument might be advanced that emotional or other strain may have prevented valid measurement on the first test, still the gain is too large and the development too marked to attribute the change solely to test conditions. The same is true of the one student in the Class of 1943 who gained sixty-six points and of the one in the Class of 1945 who gained fifty-nine percentile points. No student in the Class of 1943 lost in total points, but three in each of the other two classes lost five, eight, and ten points, respectively. The two whose ratings were decreased by five and ten points were under severe emotional strain during their college years, while two others whose total percentile rank was lower in the senior than in the freshman year appeared to be poorly adjusted to college though they did acceptable scholastic work. It is hard to explain the loss of three percentile points by another student who had an exceptionally high scholastic record except that for some unknown reason the test given in the senior year did not measure her ability accurately.

In terms of Q-score percentile points, seventy-nine, seventy-three, and seventy-one—a total of two hundred twenty-three of the two hundred eighty-eight individuals included in the study—increased their rating. The range of change in Q-score percentiles was −37 to 67, −39 to 81, and −39 to 58 for each of the three classes. About 6 per cent gained fifty or more percentile points; about eleven per cent increased their rating by forty or more percentile points; eighteen per cent gained thirty

or more percentile points; thirty-five per cent gained twenty or more; fifty-three, fifty-three, and thirty-nine individual students or 50.4 per cent of the whole number increased their Q-score rating by ten or more percentile points; 77.4 per cent of the two hundred eighty-eight students gained one or more percentile points. A gain of eighty-one Q-score points was made by a Home Economics major of the Class of 1943. These students have considerable work in chemistry which may partly explain the huge Q-score increases." When considered with reference to the total college record, however, the gains should probably be ascribed, in part at least, to emotional or other tension during the freshman test, since all three of the student's psychological ratings were low in spite of consistently good high-school and college achievement. A gain of sixty-seven Q-score points was made by a mathematics major in 1945, but a gain of sixty-two points was also made by a Spanish major, sixty-six by a sociology major, fifty-nine by a history major and fifty-seven by a psychology major, proving that in departments highly linguistic large gains in Q-score were made. Though, in general, science and mathematics majors made large increases in Q-score percentile points, it cannot be concluded, therefore, that courses involving mathematics are alone responsible for the large Q-score increases, since students who had only one year of science or mathematics made such marked changes in their quantitative ratings. Augmented mental maturity must account for many of the increases in percentile ranks. The results surely argue for the influence of training in reflective thinking on the development of quantitative mental ability.

The analysis of L-score changes showed that losses were small and few and gains large and numerous, due perhaps to the type of curricula followed by the majority of the students at Marygrove as well as to the fact that women in general rank higher in linguistic than in quantitative ability. L-score increases were made by one hundred five, eighty-two and eighty-six or a total of two hundred seventy-three students, 94.8 per cent of the entire number. Seventy-five, fifty-one, and thirty-seven—a total of one hundred sixty-three students or 56.6 per cent of the two hundred eighty-eight included in the investigation—increased their rank by thirty or more percentile points. Of the thirty-one, thirty-one, and forty-nine students who gained

between one and nineteen percentile points a large percentage had scored between the 80th and 98th percentiles as freshmen, so larger gains were impossible. Six students kept the same rating, four because they had made the 99th percentile on the first test. Nine students in the three classes lost from one to eleven L-score percentile points on the retest. The range in L-score gains and losses for the three classes was −7 to 66; −11 to 57; and −5 to 69. The greatest gain was made by a student majoring in sociology in 1946, but a gain of 66 L-score points was also made by an English major in 1943 and gains ranging from fifty to fifty-nine points were discovered in students in history, Spanish, home economics, art, and psychology. Again it must be concluded that the department in which a student does his major work is not the determining factor in the psychological change involved in the individual student. Large gains were made by science and mathematics majors in Q-score, but these majors were consistently missing from the list of those who made large L-score increases. The question arises: Is this because, since mathematics is a specific ability, these students were gifted only in this field and lacking in linguistic possibilities, or because they had reached a high degree of development along linguistic lines when the first test was taken making large increases impossible?

Size of gain alone, however, does not determine the influence of specific fields of study on the student's growth and development. As was mentioned before, large increases are impossible for persons reaching high percentiles as freshmen. On the other hand it is a matter of simple logic to see that a student who increased his first score by fifty or sixty percentile points could not have rated a high percentile rank to begin with. Since the matter is of current interest in the present widespread discussion of the value of general education, an attempt was made to determine whether or not certain types of curricula influence growth and development at the college level more than others. A method was devised that does not penalize students who reach the higher stages of development before they enter a department. Taking ninety-nine as the highest percentile attainable, the possible amount of increase in percentile points was found by subtracting the rank attained on the freshman test from ninety-nine. Thus if a student reached the 54th percentile as a fresh-

man it was possible for him to increase his rank by forty-five percentile points. The percentage of gain for each individual in the various departments was found by dividing the actual gain by the potential increase. Thus if the student in the example cited above increased his score on the senior test by twenty percentile points his percentage of gain would be the actual gain—twenty—divided by the possible gain forty-five or 44.4 per cent. Whereas considering gain alone would penalize the student attaining a high percentile rank in the freshman year, this method, in a way, favors him for, in some cases, a gain of a point or two means an increase of one hundred per cent. In spite of this defect, however, the method was considered more valid than merely considering the amount of increase in percentile points.

At Marygrove the first two years of college work may definitely be classified as general education. Since all juniors and seniors, irrespective of their field of specialization, are required to include courses in Religion, philosophy, and seminar in their junior and senior years, every student at Marygrove also continues her general education in the upper college. Each student, however, at the beginning of her junior year chooses a field for specialization. In this study certain majors have been combined to form four fields, viz., Fine Arts, Language, Science, and Social Science. Fine Arts includes majors in Music and Art; Language means

TABLE X.—PERCENTAGE OF IMPROVEMENT MADE BY STUDENTS IN THE DIFFERENT FIELDS

	Fine Arts	Social Science	Science	Language
	N-26	N-101	N-46	N-30
Q-Score	27	29	48	32
L-Score	52	58	54	72
Total	50	52	54	62

majors in English, French, Spanish, and Latin; Science is understood to include majors in Chemistry, Biology and Mathematics, while Social Science has been interpreted as including majors in History, Economics and Sociology. This classification omits from the evaluation students majoring in Home Economics, Journalism and Speech. To secure representative sampling the three classes were grouped together in this part of the study. The results are summarized in Table X.

Inspection of Table X indicates that the students specializing in Science made the highest percentage of improvement in Q-score while those majoring in the Language Arts made the greatest gain in L-score. These results are exactly as one might expect, since students having definite specific abilities in language or mathematics naturally choose departments where success will be greatest. On the other hand, it is interesting to note that all groups made comparatively large gains in Q-scores when compared with the increase of the Science group. Likewise all fields show large increases in L-score. This, too, is only what one would expect of groups that are all continuing their general education which, as explained before, is the case at Marygrove. So far as total percentage of improvement is concerned, the two groups receiving specific guidance in science and language made such large increases in their special fields that the total score necessarily exceeds that of the other groups. On the other hand, gains of all groups are relatively large. It seems that the most logical conclusion that can be drawn from this evidence is that no particular field produces noticeably greater growth than any other. The student in a liberal arts college is trained to search for truth and in the problems that he solves in this search he learns to think and to reason, irrespective of the material he uses to think with. This conclusion substantiates the evidence given by Flory[2] who, in a similar study of seventy-four students at Lawrence College, found mean gains of 7.8, 8.3, 8.0 and 5.6 raw score points for students majoring in Fine Arts, Language Arts, Science and Social Science, respectively. He concluded that none of "the large fields of study had superiority over any other large fields of study in contributing to intellectual development at the college level."

In order to discover the relation of the original status to the change in percentile ratings, the correlation between the percentile ranks in the freshman and senior years was found. The correlations are presented in Table XI. Inspection of the table proves that all correlation coefficients indicate considerable relationship. The coefficients for the total score of $.80 \pm .023$, $.82 \pm .024$, and $.90 \pm .013$ are quite like those of $.82 \pm .03$ found by Flory,[2] of McConnell[7] who found a relationship of .82 for Cornell students, and of Livesay[6] who for fifty students at the University of Hawaii found the coefficient to be $.88 \pm .021$.

For one hundred five students at the University of Illinois Barnes[1] gives the coefficient of correlation between tests taken in the freshman and sophomore years as: Q-score .75; L-score .76; and gross score .78. In the present study the average Q-score correlation is .73, the average L-score .83, while the average of the total score correlation is about .84. There is remarkable similarity in the correlation of the L-scores especially for the 1945 and 1946 classes. The correlation for the total score indicates an increase in relationship each year ranging from .80 ± .023 in 1943 to .90 ± .013 in 1946.

TABLE XI.—CORRELATION BETWEEN THE FRESHMEN AND SENIOR PERCENTILES

Class of 1943		Class of 1945		Class of 1946	
Q-Score	.67 ± .031	Q-Score	.73 ± .021	.80 ± .025	Q-Score
L-Score	.81 ± .022	L-Score	.84 ± .020	.84 ± .020	L-Score
Total	.80 ± .023	Total	.82 ± .024	.90 ± .013	Total

All seniors at Marygrove also take the Coöperative General Culture Test. As a matter of interest and to determine the degree of relationship between the percentile ranks as expressed by the psychological tests and achievement, as measured by the General Culture tests, the coefficients of correlation for the 1945 and 1946 classes were found. They were .78 ± .027 for the Class of 1946 and .71 ± .039 for the Class of 1945, both quite significant. They at least indicate that the relationship between ability and achievement, as measured by the tests used, is relatively high.

The data discovered in this investigation cannot but cause one to question the advisability of placing too much stress on test percentiles in counseling or dropping students. Only when used in conjunction with other data, such as a study of individual student achievement and personality traits, should the results be emphasized. Used with other guidance techniques they serve a definite purpose, however. As in Hartson's study, (Hartson used the Ohio State Examination at Oberlin College) it was noted that many of the students in the lowest deciles of the A. C. E. test taken in the freshman year at Marygrove left or were dropped from college, but those in the low deciles who survived for the four years showed considerable gain in percentile ranking.

Certainly the results of the present study support the conclusion of Hunter[5] that improvement of "intellectual abilities continues during the college years" provided that appropriate stimulation and constant training in thinking and in the application of knowledge are provided. The results also indicate that the rate of growth of college students differ and the consequent need of careful individual guidance in directing, especially students in the lower deciles, into majors where growth and, therefore, success is possible. They indicate the need of unwearied effort on the part of the faculty to help all students—the one whose individual development is slow as well as the more mature—to realize their potentialities and thereby reach their individual maximum level of development. The fact that Language majors gained the most on L-scores and Mathematics majors gained the most on Q-scores suggests that growth and development in these specific abilities is responsible for some change. Still, the fact that so many students who did not choose a definitely quantitative nor linguistic curriculum made so much improvement cannot but suggest, too, that training and definite functioning of the abilities needed are perhaps the most effective factors in improvement. As much opportunity as possible should be provided for the problem-solving type of procedure in college technique. Provisions should be made to give students the opportunity to undertake, plan, and carry to completion definite problems demanding the use of the higher thought processes, and this in co-curricular as well as in curricular activities.

SUMMARY AND CONCLUSIONS

The investigation here reported is a study of the changes in the psychological ratings of two hundred eighty-eight students of Marygrove College, Detroit, over a period of seven semesters of college experience. Conclusions are drawn from the analysis and interpretation of the records of one hundred seven students of the Class of 1943, eighty-six of the Class of 1945, and ninety-five of the Class of 1946, all of whom took the A. C. E. Test at the beginning of the freshman year and in the middle of the senior year. From the data obtained the following conclusions appear to be justified.

1) Highly significant gains on all scores derived from the test indicate that there is considerable intellectual growth during the

period of the college years. The mean gains in totals score in terms of percentile points for the three groups were twenty-eight, twenty-seven, and twenty-six, respectively; for the Q-scores the gains were sixteen, twenty-five, and twenty-one; and for the L-scores increases of thirty-three, twenty-eight, and twenty-three percentile points were found.

2) On the gross score two hundred seventy-seven students, or 95.8 per cent of the total number, gained percentile points ranging from one to sixty-six, one to seventy-three and one to fifty for each of the three classes. Five students in the three groups kept the same total score—three because they had reached the 99th percentile as freshmen so that no test gain was possible. Nine students, or 2.1 per cent of the three classes, lost points on the total score, but in no case was the loss especially heavy ranging from one to ten points. In terms of Q-scores seventy-nine, seventy-three, and seventy-one—a total of two hundred twenty-three students or 77.4 per cent—gained percentile points; 4 or 1.4 per cent remained the same and a total of sixty-one or 21.2 per cent for the three groups lost percentile points. The losses represented nearly every department and were not confined to those following an essentially linguistic program. Gains, too, were made by students of all departments, indicating that intellectual maturity and the influence of a mind trained to do reflective thinking were responsible for the improvement. On L-Scores gains ranging from one to seventy-three percentile points were made by one hundred five, eighty-two, and eighty-six or a total of two hundred seventy-three students—94.8 per cent of the entire number. Six kept the same L-Score and nine or three per cent of the three groups, lost in L-Score percentile points.

3) Though the students majoring in Science showed greater changes in Q-scores, and those specializing in Language attained the highest L-Scores, still the field of specialization did not affect the percentage of improvement of the individual to any great extent. Significant increases in both scores were made by nearly all students irrespective of their major subject.

4) The correlation coefficients between the freshman and senior percentiles were found to be .67 $\pm$ .031, .73 $\pm$ .021 and .80 $\pm$ .025 for Q-scores; .80 $\pm$ .023, .82 $\pm$ .024 and .90 $\pm$ .013 for gross scores; .81 $\pm$ .022, .84 $\pm$ .020 and .84 $\pm$ .020 for L-scores.

Coefficients of .71 ± .039 and .78 ± .027 were found between the percentiles reached on the Coöperative General Culture test and the psychological test administered during the senior year, indicating a fairly high relationship between ability and achievement as measured by the tests used.

5) The fact that so many of the two hundred eighty-eight students representing three college classes made statistically significant increases on the total score—78.8 per cent gaining ten or more percentile points—not only proves that college experience is highly beneficial to the mental growth of the student, but also presents a stimulating challenge to the college. The study suggests that the challenge may be met, in part at least, as follows:

(a) Curricular and co-curricular offerings should be studied constantly and changes made to provide for the mental growth of each student.

(b) The guidance of individual students into fields of study that will promote maximum growth for them as persons becomes necessary in the administration of the college.

(c) Provision for leadership should be made, not only for college entrants who have already reached a high stage of development, but for all college risks. Though slower in reaching maximum development many students in the middle, and sometimes those in the lower deciles on a psychological test, often possess immense possibilities for growth.

(d) Individual faculty members must accept the responsibility of providing for varying rates of growth and different types of development. Methods of presentation, motivation patterned to reach students of all apparent abilities, and techniques designed to develop reflective thinking, more and higher interests and appreciations, and growth in ability to work individually as well as with groups are as important in the college as in the grade and high school.

(e) When students are developing the higher thought processes of judging and reasoning, when they are solving problems significant to themselves and to society, no matter in what field of study these abilities are functioning, growth and development occur. It is the responsibility of the college to provide and organize experiences in every department of learning that will require the student to use these powers.

BIBLIOGRAPHY

1) Barnes, M. W., "Gains on the A. C. E. Psychological Examination During the Freshman-Sophomore Years." *School and Society*, 57, 1943, 250–252.

2) Flory, C. D., "The Intellectual Growth of College Students." *Journal of Educational Research*, 33, 1940, 443–441.

3) Freeman, F. N. and Flory, C. D., "Growth in Intellectual Ability as Measured by Repeated Tests." *Monographs for the Society for Research in Child Development*, II, No. 2, 1937, xi–116.

4) Hartson, L. D., "Does College Training Influence Test Intelligence?" *Journal of Educational Psychology*, 27, 1936, 481–491.

5) Hunter, E. C., "Changes in Scores of College Students on the American Council Psychological Examination at Yearly Intervals during the College Course." *Journal of Educational Research*, 36, 1942–1943, 284–291.

6) Livesay, T. M., "Does Test Intelligence Increase at the College Level?" *Journal of Educational Psychology*, 30, 1939, 63–68.

7) McConnell, T. R., "Changes in Scores on the Psychological Examination of the American Council on Education from Freshman to Senior Year." *Journal of Educational Psychology*, 25, 1934, 66–69.

8) Rogers, A. I., "The Growth of Intelligence at the College Level." *School and Society*, 31, 1930, 693–699.

9) Wright, M. B., "The Development of Mental Ability at the College-Adult Level." *Journal of Educational Psychology*, 22, 1931, 610–628.

INDIVIDUALIZED AND MEANINGFUL INSTRUCTION IN ARITHMETIC

J. W. TILTON
Yale University

I. THE VALUE

Individual remedial instruction in addition, subtraction, and multiplication was given for a short period in four fourth grades. The work was undertaken to test the belief that a very small amount of individualized instruction can produce demonstrable results.

The Compass Survey Tests in Arithmetic, Elementary Examination, Grades 2, 3, 4: Forms A and B were administered to the one hundred thirty-eight pupils in the four fourth grades. Thirty-eight of the one hundred thirty-eight pupils were selected for the experiment because they made low scores and could be paired into experimental and control groups. The average initial scores were 114.7 and 114.8 for the experimental and control groups, respectively. The standard deviations were 17.5 and 16.1 in the same order. No discrepancy was more than 2.3 times its standard error (reliability coefficient for initial scores was .93). The result was a correlation of .98 between the pairs. There were nine boys and ten girls in the experimental group, ten boys and nine girls in the control group. Since the average ages were 10-2 and 9-11, it is quite possible that the experimental group was slightly duller than the control group.

The initial tests were given late in March. During the following four weeks, twenty minutes a week were spent with each experimental pupil. Remedial effort was based on the pupils' performances on the Los Angeles Diagnostic Tests. A three-week period was allowed to elapse between the last twenty minutes of individualized teaching and the follow-up test. The follow-up test was a repetition of the Form A used as part of the initial testing. It was administered as it was seven weeks earlier to the whole grades. Practice effect, if any, was irrelevant to the problem since the gains made by the control pupils were subtracted from the gains made by the experimental pupils. The differences obtained are then freed from the

possible influence of such factors as practice effect and regression of low scores toward the mean, and are measures of the influence of the eighty minutes of individual instruction.

The experimental pupils, who were about a year retarded in addition, subtraction and multiplication skills, made five months more progress in these areas than they would have made without the individual instruction. This five-month average superiority is statistically significant at the 1 per cent level (degrees of freedom = 18). It is evident, therefore, that a small amount of individualized instruction can be very worth while.

II. THE NEED

Concerning the faulty habits involved, the facts for these pupils are, of course, in agreement with the often reported finding that difficulty is primarily a matter of not knowing the combinations and hence depending upon counting, a matter of zero difficulty, and difficulty in carrying and borrowing. There is general awareness that some of the pupils in each grade have these difficulties, and no comment is called for. The errors made by the experimental pupils, however, reveal a fact which is worth emphasizing. It is that many errors are systematic. In other words, not as many errors are accidental and attributable to carelessness as teachers are inclined to think. Take a subtraction example as an illustration:

$$\begin{array}{r} 770 \\ -545 \\ \hline 235 \end{array}$$

One would certainly think that this error was due to forgetting having borrowed, or to carelessly subtracting 4 from 6. Actually, the child who worked this example was working in accordance with her rules. The trouble was, her rules were wrong. This child was consistently taking the smaller digit from the larger one. 53 − 49 for this child was 16, 103 − 98 was 195, and 954 − 475 was 521. If the youngsters who have such incorrect

rules are to be helped, the teacher should know the child's rule, because the child's need is just as much to unlearn his incorrect rule as it is to learn the correct rule. To work in ignorance of his rules is to give him a feeling of confusion. If the teacher knows what the child's incorrect rules are, the remedial teaching is more intelligently directed to the need, and progress is greatly speeded. These remarks have been illustrated by only one example, but they might have been supported by the habit of

$$\begin{array}{r} 7 \\ +14 \\ \hline 12 \end{array} \qquad \begin{array}{r} 15 \\ +3 \\ \hline 9 \end{array}$$

adding single digits, which the teacher is in danger of interpreting as "writing down whatever answer comes into his head;" or

$$\begin{array}{r} 760 \\ -15 \\ \hline 655 \end{array}$$

in which case the child takes the smaller digit from the larger one over or under it so far as they are paired and then pairs the odd digit not already used, with one used before, (in this case 7 and 1); or

$$\begin{array}{r} 32 \\ \times 49 \\ \hline 88 \end{array}$$

illustrating the hybrid habit of multiplying, carrying and adding. The teacher is not apt to know of these incorrect rules in accordance with which children work unless the teacher takes time enough with the individual child to see, if there is a consistency in error, what the habit is.

Many of the errors made by these fourth-grade children seemed to be due to an insufficient understanding of the meaning of numbers. It seems as if these children had been asked to learn the rules for the manipulation of numbers in addition, subtraction and multiplication without having learned the meaning of the symbols they have been asked to manipulate. Because so many children learn a great deal of the meaning of the number system without the teacher's assistance the teacher is apt to assume such understanding. But the glib verbalization of numbers in their proper order is no guarantee

that the system of symbols for units, tens, etc., has made sufficiently meaningful in terms of the child's concrete experience. Probably such a foundation is sufficiently laid in the case of some children.

Possibly all children might, in the long run, profit from more concrete familiarity with numbers. After an adequate experience of this sort a child would not add the separate digits $\begin{array}{r} 7 \\ +14 \\ \hline 12 \end{array}$. The common habit of taking the smaller digit from the larger digit would not have developed, and a child could not be satisfied with $\begin{array}{r} 84 \\ \times 10 \\ \hline 80 \end{array}$ obtained by multiplying the vertically paired digits, 0×4, and 1×8. Knowing what 45 and 26 mean, a child faced with $\begin{array}{r} 45 \\ -26 \\ \hline \end{array}$ wouldn't ask, "Which way do you go?", not knowing whether to take the upper from the lower or vice versa. Nor would a child who, in a mechanical way, knows how to borrow, ask, "Is there any borrowing in this?" Nor would another child have trouble remembering which digit to carry and which one to put down. Carrying and borrowing can be taught quickly to some of the children who have been slow to learn them by making the processes meaningful. One child had a rule for carrying which made some of her answers right and some wrong. She carried 'the big one,' i.e., the larger of the two digits. In other words, for her, $\begin{array}{r} 46 \\ \times 3 \\ \hline 201 \end{array}$ and $\begin{array}{r} 47 \\ \times 3 \\ \hline 141 \end{array}$.

It is quite possible that habits like this one have their origin in instruction misinterpreted. For instance, a teacher might speak of the tens digit as 'the big one' and the units digit as 'the little one' and tell the pupils to put down the little one and carry the big one. But, for a child who doesn't understand

numbers, who knows only digits, the instruction would become an incorrect rule. In a similar way, a teacher might say of the numbers in subtraction examples, "we always take the little one from the big one." But again, for the child who sees in numbers only digits, the teacher's remark about numbers is understood as referring to digits. I do not know that such teacher remarks are responsible for the habit of always subtracting the smaller digit from the larger digit, but as teachers we should be constantly critical of our ways of saying things in an effort to avoid the inadvertant teaching of error. Perhaps we make subtraction hard for children by referring to $\begin{array}{r} 45 \\ -26 \\ \hline \end{array}$ sometimes as "45, take away 26," and at the other times as "subtract 26 from 45." Consistency, so long as any of the pupils are in the early stages of the process, might reduce confusion, but the most profitable preventive measure is, with plenty of practice in concrete situations to make sure that each child understands numbers.

In the meantime, i.e., until instruction is sound so far as it goes for all pupils, there is the common situation in which a few pupils in each grade are badly in need of a more basic teaching than they are likely to get in their grade. Among fourth-grade pupils developing division skills, pupils making errors such as have been illustrated in this article are not likely to get help. How can they be helped except on the basis of the diagnostic observation of the individual pupil while he is working?

III. IN CONCLUSION

It must not be inferred that needs such as have been described can be adequately met by eighty minutes of individualized instruction. At the end of the experimental period, the majority of the nineteen pupils were still retarded and handicapped for the work of the grade. The short period was chosen for evaluation to show that even a small amount of such instruction helps. In the second place, there is no pretense of originality in offering this description of the need. It is submitted rather for an emphasis which is believed desirable.

REFERENCES

Brownell, W. A., "When is Arithmetic Meaningful?" *Journal of Educational Research.* March, 1945. 38: 481–498.

Morton, J. A., "Developing an Understanding of Our Number System Through Place Value." *School Science and Mathematics.* December, 1945. 45: 870–873.

Riess, Anita, "Meaning of the Meaningful Teaching of Arithmetic." *Elementary School Journal*, September, 1944. 45:23–32.

ACTIVE VERSUS BLANK RESPONSES TO MULTIPLE-CHOICE ITEMS

ALEXANDER G. WESMAN

The Psychological Corporation

In the process of constructing multiple-choice tests, one is sometimes forced to decide whether to require the subject to mark every possible response or whether to require him to mark only choices of a given category. For example, we may have a spelling test in which the subject is required to mark only those words which have been wrongly spelled, or we may require him to mark every word as being correct or incorrect. In the first instance, we consider all blank responses as indicating the subject's judgment that these words are properly spelled. In the second instance, we evaluate only his active responses.

The blank response form is probably to be preferred on the basis of the time factor; more items can be answered in a given period of time when only some items require an active response than when all items require marking. The speed advantage will vary according to the complexity of response required and test format. If the marking is simple, the advantage will be relatively small; whereas, if the marking is complicated, the advantage of the blank response form may be appreciable.

The blank response form has at least one disadvantage which may outweigh the advantage of its permitting greater speed. This is the confusion which frequently results from our inability to know whether the blank response really represents a judgment by the subject, or whether it is merely an 'omit.' This difficulty is absent only in a pure 'power' test, when everyone has all the time he needs for every item. To the extent that speed is a factor in the test performance, the confusion element becomes increasingly important. If a subject does not finish the test, it is difficult to assign an equitable score. For example, let us suppose that the spelling test mentioned above has fifty words and the subject has marked the fortieth word as being incorrect, leaving the last ten words unmarked. Unless we are sure he has had enough time to finish the test, we do not know whether he looked at any or all of the last ten words and considered them as being properly spelled. If we assign a score on the basis of the

first forty words only, and he has really made judgments about some or all of the remaining ten words, he is being penalized unfairly if more than five of the last ten words are truly correctly spelled, or not being penalized adequately if fewer than five are correctly spelled. Similarly, if we assign his score on the basis of all fifty items, and he has actually not looked past the fortieth item, he is penalized unfairly or rewarded unfairly depending on the number of correctly spelled words among the last ten items of the test.

Theoretically, we may expect that, since the active response form eliminates the ambiguity of the blank response form, there should result a greater reliability. The study reported here is an investigation of the validity of this hypothesis.

The test used consisted of fifty sentences, each sentence being divided into five parts (items). The subject was required to decide whether each part of each sentence was grammatically correct, or whether one or more of the parts were wrong. Two forms of the test were available. They were administered to two hundred ninety-nine ninth-grade high-school students.* Form A was administered first, with instructions to one hundred twenty-nine of the students (Group I) to mark every part 'Right' or 'Wrong' on a special answer sheet. The one hundred seventy students in Group II were instructed only to pick out those sentence parts which were 'Wrong,' and mark those responses on an appropriate answer sheet. Form B was then administered to these two groups, but with reversed directions: Group I now was instructed to mark 'Wrong' parts only; Group II marked every sentence part as 'Right' or 'Wrong.'

The per cent passing each item (each sentence part) was found. Since there were fifty sentences and five parts to a sentence in each form, it was necessary to compute 1000 per cent passing values. This number is too large to present for meaningful comparisons. Accordingly, the five per cents for each sentence for each form were averaged. The average per cent passing was: Form A, active response, sixty-seven per cent; Form A, blank response, seventy per cent; Form B, active response, sixty per cent; Form B, blank response, seventy-eight per cent.

---

* The author is indebted to Leslie O. Johnson, Principal, The High School, Gloucester, Massachusetts, for the testing of these students.

On the whole, the items seem to be a little more difficult under the active response condition; i.e., when the subject is asked to mark both 'rights' and 'wrongs.' The greater difference in mean per cent passing on Form B than on Form A may result from a combination of the following: 1) there is a real difference in difficulty between the two forms;* 2) the transition from marking 'wrongs' only to marking both 'rights' and 'wrongs,' and *vice versa*, may result in negative transfer in marking Form B, which was the second test taken. As for the difference in means within each form under the two experimental conditions, the active form of response may actually increase the difficulty of the items; or the greater time required by the active form may prevent more subjects from reaching the last items of the test, and thus serve artificially to make those items appear more difficult. To investigate the latter hypothesis, distributions were made of last item attempted (as judged by positive marks on the answer sheets). The data are presented in Table I, and reveal that the active

TABLE I.—LAST ITEM COMPLETED UNDER ACTIVE AND BLANK RESPONSE CONDITIONS

Last Item Completed	Form A		Form B	
	Active Response	Blank Response	Active Response	Blank Response
49–50	28%	56%	21%	76%
40–48	39%	21%	34%	19%
Below 40	33%	23%	45%	5%
N	170	129	136	165

response was indeed more time-consuming and actually did not permit as many subjects to reach the last items of the test. As further vertification, the average per cent passing was computed for the first twenty-five and last twenty-five sentences separately. Table II presents these means, and confirms the fact that it is the failure of subjects to reach the last items that is responsible for the smaller per cent passing on the active form of the test. The

* The author has evidence from another study that, by roughly twelve points in score, Form B is actually somewhat more difficult than Form A.

differences in means are much larger for the second half of the test than for the first.

'Per cent passing' in this study was calculated as the proportion of the entire group taking the test which answered an item correctly. The observed per cents would increase if only those actually reacting to each item were used as the base for computation. This analysis was not attempted because it would require guessing as to whether, under blank response conditions, the subjects had looked at one or more items after the last one marked and found no errors in grammar, or whether failure to mark the items was truly an indication that those items had not been read.

TABLE II.—AVERAGE PER CENT PASSING FIRST AND SECOND TWENTY-FIVE SENTENCES UNDER ACTIVE AND BLANK RESPONSE CONDITIONS

	Form A		Form B	
	Active Response	Blank Response	Active Response	Blank Response
Sentences 1–25.....	83%	78%	79%	86%
Sentences 26–50.....	51%	62%	40%	71%

Item-test correlation coefficients were computed for Forms A and B of the test under active and blank response conditions by use of upper and lower twenty-seven per cent groups and the Flanagan chart.* The average item-test coefficients were: Form A, active response, .41; Form A, blank response, .25; Form B, active response, .35; Form B, blank response, .25. With both forms, the average item-test r is higher under the active response condition than under the blank response condition.

It would be possible for this superiority of the active response condition to be the result of the same factor which caused apparent increased difficulty; namely, the relatively few subjects answering the last items of the test. This hypothesis can be tested. It is presumed that some imperfect but positive relation-

* Flanagan, J. C. *A table of the values of the product moment coefficient of correlation in a normal bivariate population corresponding to given proportions of successes.* New York: Coöperative Test Service, 1936.

ship exists between the subject's competence in the content of the test and his speed of response to the items; i.e., there is a tendency for the better subjects to answer more items. Therefore, these better subjects will reach and correctly answer some of the items near the end of the test, while poorer subjects (many of whom might have answered the items correctly) do not reach these items. To the degree that these items are easy, the result is artificially high item-test coefficients.

To discover whether this would account for the result cited above, average item-test coefficients were computed for the first twenty-five and last twenty-five sentences separately. The data are presented in Table III, and reveal that there is almost

TABLE III.—AVERAGE ITEM-TEST COEFFICIENTS FOR FIRST AND SECOND TWENTY-FIVE SENTENCES UNDER ACTIVE AND BLANK RESPONSE CONDITIONS

	Form A		Form B	
	Active Response	Blank Response	Active Response	Blank Response
Sentences 1–25.....	.27	.25	.33	.24
Sentences 26–50.....	.55	.25	.37	.26

certainly an artificial character in the item-test r's for the last items under the active response condition. Nevertheless, the average item-test r is higher even for the first twenty-five sentences under the active response condition. Apparently, requiring active responses does improve item-test correlation somewhat.

What effect has the use of one or the other type of response on the reliability and validity of the test? The strong probability is that validity will not be affected, except insofar as validity is dependent on reliability. At least the content of the test is the same under both active and blank response conditions. No outside criterion was available which would permit investigation of this question. The reliability of the test, on the other hand, might be expected to vary with the type of response used. Since more subjects answer a larger number of items where they are asked to mark wrong items only, this blank response condition

results effectively in a longer test than the active response situation. This greater effective length of test makes for additional reliability. However, the individual items should be more reliable when active responses are required for all items, and this greater individual item reliability should summate to increased test reliability. Which force (if either) is the more effective in the experimental situation?

Coefficients of correlation between odd and even halves of each form of the test under each response condition were computed for boys and girls separately and combined. The corrected (by Spearman-Brown formula) coefficients are shown in Table IV.

TABLE IV*.—RELIABILITY COEFFICIENTS AND CRITICAL RATIOS UNDER ACTIVE AND BLANK RESPONSE CONDITIONS

	Form A						Form B					
	Boys		Girls		Boys and Girls		Boys		Girls		Boys and Girls	
	r_{11}	N	r_{11}	N	r_{11}	N	r_{11}	N	r_{11}	N	r_{11}	N
Active.....	.88	53	.93	117	.92	170	.88	91	.90	45	.89	136
Blank.....	.86	83	.89	46	.86	129	.86	55	.91	110	.89	165
CR.....	.45		1.32		2.50		.48		.30		0	

* The reliability coefficients are the odd-even correlations stepped up by the Spearman-Brown formula. The CR was obtained by converting r to Fisher's z-function and getting ratio of difference between z's to standard error of difference.

These coefficients are somewhat spuriously high of course, since the tests had an element of speed involved; and since speed was a more important factor under the active response condition, the coefficients are probably more spurious for that response than for the blank condition. Critical ratios of differences in the coefficients were computed, using the Fisher z function.

There appears to be no real difference, in reliability, between the active and blank conditions. The one critical ratio which approaches statistical significance (2.50 for total group, Form A)

applies to an instance in which the active response coefficient is higher. Since it is only one difference out of six, and since (as indicated above) there is more spurious magnification in the active response test coefficient, the evidence can only be accepted as showing no real difference.

If the results of this experiment are typical for situations in which one is faced with the choice of using active or blank responses, the choice would seemingly depend on mechanical advantages in a specific situation rather than on any inherent superiority of measurement. For equal testing time the theoretical advantages of greater sampling under the blank condition, and lesser ambiguity of interpretation (and consequent item reliability) of the active condition, seem to equal each other. The appropriate considerations for reaching a decision in a specific situation will probably depend on such factors as the desired magnitude of range in raw scores, the applicability of one from or the other to mechanical aids such as standard IBM answer sheets, the relative difficulty of writing simple instructions for the population for which the test is intended, the scoring problem, etc.

The writer is aware that it is dangerous to generalize from a single experiment. Variations in population and test content might produce different results from those reported herein. The present study does serve, however, to demonstrate the greater danger of making decisions regarding item types on the basis of armchair logic, and the need of exposing hypotheses concerning items to experimental investigation.

DRAMATIC DIALOGUES FOR SIMULTANEOUS TREATMENT OF READING AND PERSONALITY PROBLEMS

MARY McGANN

Worcester, Mass.

Reading and personality problems are so closely interwoven that progress in either is often a reciprocal process. Rapid recovery requires materials for sufficient treatment of weaknesses in both reading and personality. For this purpose, dramatic dialogues written to suit the needs and interests of the individual, have been found to have unique values.

ORIGIN OF DRAMATIC DIALOGUES

The author, in efforts to capture the interest of an overactive, aggressive, very superior sixth-grade girl with third-grade reading ability, observed that the reading of dramatic material had distinct therapeutic effects. Suitable dramatic material, however, was not available in sufficient quantities for a series of reading lessons. Even the simplest plays were complicated by interpolated, uninteresting content: stage directions, descriptions of characters, and scenery. The vocabulary, too, was difficult.

It was obvious that most of the dramatic material available was written to be presented on the stage. A search for more simplified plays revealed that such materials did not exist. The closest approximation was the radio script, but that was seldom published, and even if it were, the content would not be suitable.

To satisfy the specific needs of the pupil, the author devised her own dramatic dialogues. Since they proved to be satisfactory, many other dialogues were written to fill the requirements of the other retarded readers under her instruction.

A detailed study[1] of the dramatic dialogues used in the recovery of these pupils was recently made by the author at the Psycho-Educational Clinic of the Harvard Graduate School of Education

[1] McGann, M. E. *A Study of Dramatic Dialogues Adapted for Remediation in Reading.* Cambridge: Harvard Graduate School of Education, Ed. D. Thesis, 1946.

under the direction of Dr. Walter F. Dearborn. Seven case studies were presented, together with samples of dramatic dialogues used in the instruction. The quoted reactions of many other teachers of remedial reading who had an opportunity to use the dramatic dialogues for the treatment of similar cases expressed unanimous approval, together with the recognition of many psychological values inherent in the use of the dialogue materials.

DESCRIPTION OF DRAMATIC DIALOGUES

The dramatic dialogues for remedial reading were extremely simplified in order to retain the advantages and eliminate the disadvantages of using dramatic material for reading.

There were only two characters, permitting the pupil and teacher to maintain an unbroken conversation. Descriptions of characters, stage directions, and scenery were not given. A change in characters or in setting was indicated by a space containing a series of asterisks. This proved to be self-explanatory, and entirely satisfactory for the purpose. The prime requisite of the dramatic dialogues was good conversation. The content was determined by the interests and needs of the readers.

At the beginning of the dialogue was a list of the more difficult words for preliminary study. At the conclusions, a series of ten multiple-choice questions were included as a check on comprehension.

PUPIL REACTIONS

Ruth, the very superior, restless girl with whom the dialogue materials originated, would ask to have them read at the beginning of the instruction, saying, "We can work hard afterwards."[2] She would giggle in glee at the subtle wit in dramatic dialogues from *Alice in Wonderland*, as:

Hare: No room! No room! I say there is no room for you here!

Alice: There is plenty of room! [The Dormouse looks very uncomfortable.]

Hare: Have some wine!

[2] Ibid. Chapter III, Case A.

Alice: I do not see any wine. There is nothing on the table but tea.
Hare: There is no wine.
Alice: It was very polite of you to offer it, then.
Hare: It was not very polite of you to sit down without being invited.[3]

Ruth's problem in reading had been caused by an over-ambitious mother who had taught her to read before she started school in the first grade at the age of four. Ruth's need for aggression is provided for in the above type of dialogue, as well as in the following sample, where she takes the part of the child:

Bobbie: I don't want any milk. I don't want it.
Mother: But you need it, Bobbie. You are getting so thin that the wind can blow you away.
Bobbie: I like the wind. I don't like milk. Let the wind blow me away. See if I care.
Mother: But I care, Bobbie. I would never see you again—[4]

Humorous dialogues, too, helped to secure rapport and reduce tension prevalent among the retarded readers, at the same time permitting the attitude of pleasure to supplant the usual aversion toward reading:

Question: What is the best way to make a coat last?
Mark: Make the vest and trousers first.
Question: Why is a half moon heavier than a full moon?
Mark: Because a full moon is lighter.[5]

Michael (Case D in the third chapter of the study on which this article is based) was a very superior eleven-year-old boy with a severe reading disability presumably due to aniseikonia. Until the third grade, no special efforts were made to treat his disability, and when he failed to be promoted, the parents became very much concerned, securing the proper eye-treatments and daily individual remedial reading lessons by the regular classroom teacher until Michael showed some evidence of possibilities for success in learning. In addition to visual defect, poor general health, with frequent absences from school

[3,4,5] Ibid.

in the early grades, had also impeded favorable progress in reading. Michael showed evidence of poor motor coördination as well as mixed dominance. His personality was characterized by extreme indifference to his failures, but when he agreed to forfeit a summer at camp in order to have daily lessons at the clinic, he revealed that he was deeply concerned with his disability. Intelligence tests showed him to be very superior, with 134 IQ on the Stanford Binet on initial testing, although his school record at the lower levels had a score of 112 on the same test. He evidently had been improving under intelligent guidance, and overcoming his many handicaps gradually.

Testing revealed that Michael's reading was between second- and third-grade level—far too low for success in sixth-grade work. He read not word by word, but letter by letter, due to out-moded methods of teaching, and when he tried to write, he showed inability to recognize sounds as well as inability to make the correct symbol, for the o, a, m, n, ve, w, etc., were confused.

His treatment required kinaesthetic methods, but his resistance was a constant problem, especially when writing was necessary. In order to arouse interest in reading, dialogue materials were written, emphasizing his interests in humor, dogs, sports, and aviation. At first he accepted the dialogues with no more than an inquiry as to where we got 'those stories.' During the third lesson, Michael had relaxed enough to forget himself and laugh out loud, "Oh, brother!" he exclaimed. "Have you any more of these stories? I love to read them!" Interest was awakened, and favorable attitudes continued to be learned.

Reading which started at the primer level was up to fifth grade before the vacation ended, for Michael managed to read on an average three easy books per week. On the Stanford Achievement Test, the average reading level was fifth grade, and the following school term, Michael continued his lessons in order to be prepared for the extensive study-type reading required of seventh-grade pupils. Without attention to the negative attitudes his progress would never have been so successful. The dialogue reading had reawakened enthusiasm to such an extent that he would select books of his own accord instead of half-heartedly agreeing to read whatever the instructor would present as interesting to him. Michael had never read a book

of his own choice before, but as he started on his vacation, he went to the library to get a book on aviation, adult in interest level, but well-illustrated, for he wished to know more about the dramatic incidents presented in his dialogue reading.

QUOTED REACTIONS OF OTHER TEACHERS OF REMEDIAL READING

My retarded reader liked the story.

He was interested.

He wanted to imitate my voice.

He was ready to read his part of the dialogue before I had hardly finished my part.

He remembered the difficult words in my part, so that when they appeared in his part he knew them.

He was able to answer all the questions, for the dialogue material motivates the reading to a great degree.

A sense of security was manifest.

I like to have teachers read to children. Some children do not ever hear good reading—just as a beginner at piano should hear beautiful music . . . the IDEAL is presented.

The novelty appealed to the subject.

The comprehension was better than usual.

. . . effect of having a partner seemed stimulating.

It provides a special opportunity for the teacher to show the pupil naturally how he should read orally and in a life-like way, rather than to criticize him as he reads alone.

The first comment by my pupil after finishing the reading: "Just like a play. I like to read plays."

The fact that two people are involved, in flesh and blood, adds zest, challenge, some imitation, and possibly a spirit of competition.

He copied some of the riddles, took them home, and reported that he had much fun trying them out on his family.

Dialogue materials seem to have the third dimensional appeal of the drama.

The selection made my pupil want to go on reading about the same people and stories.

It gives him a feeling of confidence to be reading with the instructor.

PSYCHOLOGICAL EVALUATION

Further analysis revealed some of these factors which helped to make dramatic dialogues effective in the treatment of the co-existing personality and reading problems under consideration:

1) The dialogue technique appeals to natural interest in drama and imitation. In the rôle of another character, self-consciousness and consequent tension, characteristic of disabled readers, are soon modified.

In some cases true joy in reading is experienced for the first time, with the result that prevalent antagonistic attitudes begin to weaken. Thought-consciousness replaces word-consciousness. Opportunity for vicarious experience intensifies meaning and attention.

2) The presentation of a play is time-consuming and expensive, but the reading of dramatic dialogue materials permits some of the same beneficial effects. Psychologists for many years have made use of dramatic interest in assisting the recovery of various personality problems. Such dramatic aids are evident in play therapy, psychodrama, relationship therapy, psychoanalysis, and even in the person-to-person relationship of interviews.

3) Dramatic dialogues utilize the most natural form of expression, conversation. Since earliest language experiences are oral and conversation predominates, the dialogues provide a natural transition to reading. Context, in conversational reading, acts as a stimulus to word recognition. "When words are associated in context there is a feeling of expectancy, curiosity, strain, forward push."[6]

4) Conversations predominate in favorite comics, radio and movie programs. Reading dramatic dialogues may be associated with these pleasant experiences. As in comics, the reader is not overburdened with lengthy paragraphs of solid reading.

Both action and conversation produce vivid appeal, captivating the attention so that voluntary concentration is not

---

[6] Huey, E. B. *The Psychology and Pedagogy of Reading.* New York: Macmillan Company, 1908, 143.

required, thus permitting freedom to think of meaning, a requisite of reading skill. Even dull, factual matter can be given interest through dialogue treatment.

5) There is social stimulation in sharing the reading of dialogue. It has the signal value of sustaining the urge to measure up to the partner's superior standards, motivating excellence of performance.

6) Furthermore, the novelty of the dramatic dialogues attracts immediate appeal. Other remedial reading materials retain the usual textbook flavor, recalling unhappy experiences. Fresh motivation—the changing of the set of mind or intention of the learner is a more important factor in remedial reading than the particular method of instruction.[7]

7) Dramatic dialogues satisfy the personality needs of retarded readers, restoring security and self-confidence, facilitating rapport, motivating through content, experience of success, and knowledge of progress. Provision of verbalization, projection, identification, and such opportunities for directed self-expression may facilitate the dissolution of conflicts. In this area the dramatic dialogue materials appear to be of greatest value.

8) The dialogue materials enrich reading interests. They provide the motivation for reading the complete volume from which a dramatic dialogue is sometimes adapted.

9) Dramatic dialogues provide necessary individualization. No two cases of reading disability are exactly alike, and in many cases suitable materials must be originated.

10) Dramatic dialogues can be used for both individual and group instruction. They originated with individual treatment, but experience has proved that imitation radio broadcasts, reading of dialogues for making a recording, and other group activities in oral reading with dialogues can be developed. The better reader can take the part of the teacher.

11) Improvement in oral reading is facilitated through dialogue reading. Word by word and other detrimental habits respond to the presence of a model reader for the pupil to imitate.

---

[7] Dearborn, W. F. "Motivation versus 'Control' in Remedial Reading," *Education*, LVIII (September, 1938), 1.

The teacher no longer listens to correct errors, but lightens the task of reading by taking an active part.

12) Comprehension tests at the conclusion of the dialogue reading stimulate the development of some of the numerous reading skills. They motivate thought-getting.

All the previously listed factors contribute toward the formation of an ideal learning situation for remedial instruction in reading. Experience with continuous failure has the inhibiting effect of punishment. Experimental evidence reveals that an infra-human animal, receiving punishment in place of reward or reduction of drive for its activity, will react with evidence of conflict, then go "to the opposite end of the apparatus where it indefinitely crouches."[8] There is loss of exploratory, self-initiated activity necessary for learning. Experience with a pupil's failure to read has similarly resulted in apathetic or negative attitudes toward all forms of reading activity, as well as teachers, school and books. Although the intelligence is normal or superior, there is evidence of emotional disturbance in most cases of reading disability. Until personality factors respond to treatment real progress in reading is uncertain.

SUMMARY AND CONCLUSIONS

Dramatic dialogues which treat personality problems simultaneously with reading problems have recently been developed at the Psycho-Educational Clinic of the Harvard Graduate School of Education. A detailed study of these materials used in the treatment of disabled readers by the author and other teachers of remedial reading reveals many psychological advantages.

The outstanding value of reading dialogue materials is their therapeutic effect on the personality needs of the disabled reader. The simultaneous treatment of reading disability provides motivation for effective learning through highly individualized content and novelty of method. Conversation, self-forgetfulness, social stimulation, oral reading, rapport, and the learning of favorable attitudes toward reading are some of the many advantages inherent in the use of dramatic

[8] Mowrer, O. H. *Basic Sciences of Personality.* Cambridge: Harvard Graduate School of Education. Mimeographed copy, Chapter 3, 38–39.

dialogues for remedial reading. Indeed, the materials seem to be ideal for the clinical treatment of severely disabled readers.

The development of dramatic dialogues deserves serious consideration when remedial reading materials are still lacking in sufficient quantity and quality. Most teachers of remedial reading will agree that "the dearth of material accounts, in part, for the limited reading of retarded children, and presents one of the greatest obstacles to effective therapy."[9]

---

[9] Witty, P. and Kopel, D. "Motivating Remedial Reading: The Interest Factor," *Educational Administration and Supervision*, XXII (January, 1936), 6.

SOME EFFECTS OF RADIO-LISTENING ON THE EFFICIENCY OF READING-TYPE STUDY ACTIVITIES

LEO R. MILLER

Principal, Thacher School, Kansas City, Missouri

While working as a teacher in various schools in the Kansas City system, it was my opportunity to counsel many elementary-school pupils who were planning shortly to enter high school as freshmen. In the conversations there often arose the question of demands for home-study and of family coöperation in giving the students proper places and quiet for home-work. Frequently the problem of radio interference was brought up. Most of the pupils' families had radios and, in addition, many of the boys and girls had radios in their own rooms. Some of the pupils felt that listening to radio interferred with their ability to concentrate and said they would try to secure the coöperation of their families in agreeing to a schedule under which they might have sufficient uninterrupted time to prepare assignments.

But it was interesting to find, from these talks, that about half the pupils not only felt that radio-listening did not detract from their ability to concentrate on study, but that concentration was materially helped if a radio was playing while they worked. This attitude was in such direct contrast to that of the other group of pupils that the author's curiosity was aroused. Wide ranges of intelligence and achievement were represented in both groups, as were ages, economic background, and size of families. It was decided, therefore, to conduct an experiment to determine, as nearly as possible, which group was right or more nearly right.

The purpose of this study, then, was to determine experimentally the degree to which radio-listening interferes with or stimulates the ability to concentrate on study materials.

ORGANIZATION AND PROCEDURE OF THE STUDY

The subjects for the experiment were the members of the sixth and seventh grades of the Jefferson Elementary School of Kansas City, Missouri. In the group were thirty-nine boys and forty-six girls. They were in two separate classrooms; and, except for 'catch-up' tests made necessary by absences, all tests

were given in the classrooms, under as nearly normal conditions as were possible to arrange.

Early in the month of October, all pupils were adiministered one form of the reading section of the Stanford Achievement Test, consisting of a test of ability to grasp the meaning of paragraphs and a test of word meaning or vocabulary level. The test was given under normal classroom conditions. The pupils were not told at this time the purpose of the test but they assumed that it was merely another of the many achievement or diagnostic tests given in the Kansas City system.

At the conclusion of the test the boys and girls were told of the proposed study and their coöperation was solicited. In the ensuing discussion, it was plain that there was about a 50–50 division between the attitude that radio-listening interferred with study and the attitude that it helped in studying.

During the following five weeks, different forms of the Stanford Reading Test were administered to the same subjects on five separate occasions. The only difference was that in each of these five administrations, a radio was playing in the classroom at what we chose to call 'classroom volume.' For the purpose of this study, 'classroom volume' meant that the radio amplification was increased until every boy and girl in the room could hear the program distinctly.

These five tests will be called throughout the study the 'radio tests,' numbers One to five. They were administered on days quite similar to that on which the first 'non-radio' test was given. The month of October was selected because it usually provides many consecutive days of uniform temperature and bright sunshine.

The programs which were used for this experiment came from the same station at all times; and, as nearly as possible, at the same time of day. They were afternoon broadcasts of the type called 'sustaining programs.' They consisted of both music and talks. One week after the administration of Radio Test No. 5, another non-radio test was given. In this study the term 'non-radio test' applies to the initial test given before any of the radio tests and to a follow-up test given as a check on the practice factor. There were given, in all, seven tests, in this order: Non-Radio Test 1. Radio Test 1. Radio Test 2. Radio Test 3. Radio Test 4. Radio Test 5. Non-Radio Test 2.

After the last of the seven tests was given to the group, the writer scored each test and the grade equivalents were converted into equated scores by using the table prepared by the authors of the test. The results were then divided on the basis of sex, and preliminary tables prepared, one for boys and one for girls. These data were then ready for statistical analysis.

Throughout the experiment, insofar as could be determined, interest was keen and participation wholehearted. There is no doubt that there was a lag on the part of some of the students, but it would seem that such a lag would be inevitable.

PRESENTATION OF THE DATA

As stated earlier, it was the intention of the experimenter to keep separate the data concerning boys from those concerning girls. This was done in order to determine if sex difference had any effect upon the degree to which radio-listening aided or impeded ability to study. Therefore, all the data in the report show the results of boys and of girls separately.

Figures I and II permit a quick appraisal of the experiment's results. The graphs present, first of all, the differences in total scores of the various tests as well as the part scores, shown at the bottom of the figures. Total scores are the sum of the score made on paragraph comprehension on the Stanford Achievement Test plus the score made on the word meaning section of the same test. In the lower section these part scores are shown separately. This is the only place in the discussion of results in which the separate part scores are presented. All the other data deal with total scores, since the totals probably represent a more reliable measure of study-type activity than the part scores.

In Fig. I—Boys the upper section shows that both Quartile lines as well as the median line have a fairly sharp inclination from the NR. 1 Test to the R. 1 Test. How much of this is due to the practice factor; how much, if any, to the presence of radio; or how much to quickened interest deriving from knowledge on the part of the subjects that they were participants in an experiment, is impossible to say. Most of the pupils had been tested many times and another test should have been of no particular concern to them. However, the influx of war-workers' children, principally from small town and rural schools, may have provided enough subjects not used to standardized tests that the practice

Figure I - Boys - 39 Subjects

on only one such test, the NR. 1, would have provided as advantage in taking R. 1 a few days later. This factor is more or less discounted by the graph which shows that substantial drops occur between later radio tests. If the practice factor were the only one involved, we could expect a fairly steady increase in each radio test score over the preceding one, at least until 'the point of diminishing return' was reached. But another factor could reasonably be expected to enter. That is, boredom or lack of interest as the tests progressed. We could expect this factor to balance somewhat that of practice.

As can be observed, with few exceptions R. 2 is the high point of each line. It may be that this signifies the extreme limit of high interest in the project after which boredom entered to a greater extent. It may be that at this point the effects of radio-listening while studying, began to show up significantly.

In any case the lines show fairly sharp declinations between R. 2 and R. 3. It is difficult to account for this. The two testing days were quite similar—warm and clear. There was no heat in the rooms other than atmospheric temperature since all windows were open. The two day's programs had been quite similar. They were heard at the same time of day, had the same entertainers, and were similar in content. It is possible that the content of R. 3 was more appealing or disconcerting than that of R. 2, but the experimenter was unable at the time to note any significant difference.

With only two exceptions, all lines show slight rises between R. 3 and R. 4. In only one instance, however—that of the Median, Paragraph Meaning—does the R. 4 line approach or surpass the R. 1 or R. 2 level. Beyond R. 3 there is a tendency for most of the lines to level off, to form plateaus. There are some obvious exceptions, but the tendency is rather marked.

In summary, all total score lines for the boys rise from their initial point rather sharply, continue this rise to a high point at R. 2, suffer a sharp fall to R. 3 and show a tendency to level off at NR. 2, slight rises occur. With some exceptions, this tendency exists for part scores for boys.

Fig. II—Girls follows the same general pattern of Fig. I—Boys with a few rather striking differences. With only three exceptions the lines of Fig. II show a slight decrease from NR. 1 to R. 1, while in the case of the boys all similar lines rose. As in the

Figure II-- Girls--46 Subjects

case of the boys, all lines show inclinations, some slight, some sharp, between R. 1 and R. 2. In almost every case, similarly, R. 2 represents the peak of achievement on the tests. Then, as with the boys, the scores, of the girls show a comparatively sharp drop to R. 3. In practically every case, the point reached in R. 3 is below R. 1 as well as R. 2.

The girls, in most cases, did not make as substantial recovery between R. 3 and R. 4 as did the boys. However, with the exception of the Q. 1 lines in both total and part score graphs, the tendency to level off after the abrupt drop to R. 3 is as apparent with the girls as with the boys. It seems that the boys show a slightly more pronounced tendency to recover on NR. 2 than do the girls. The girls show a slightly narrower range at most test points than do the boys. The quartile lines are separated by fewer points.

In general, the two part score lines for both boys and girls follow each other fairly well, Differences are slight. The more apparent exceptions are at R. 2, Q. 3 part score on Fig —Boys, and at R. 4, Q. 1 part score on Fig. II—Girls.

In summary, for the girls' graph it would seem that after a slight fall from NR. 1 to R. 1, there is, in general, a marked rise to R. 2, which, in every case, marks the high point in test results. After this there is a very marked declination to R. 3, followed by a leveling—off of most lines and, in some cases, a slight fall between R. 5 and NR. 2.

The lower part of Fig. I shows that there is a rise, either slight or sharp, in the quartile lines and the median lines for both paragraph comprehension and word meaning. Furthermore, both sections of the graph show that, with only four exceptions, all lines, both quartile and median of total and part scores, carry this rise on to R. 2. These four exceptions are found in: Q. 1 of total score, which drops approximately one point; Q. 3 of word meaning, which drops approximately two points; Q. 1 of word meaning, which drops approximately one point; and Q .1 of paragraph meaning, which drops approximately one point.

The majority of the data presented in Tables I and II are self-explanatory. The more significant parts are those dealing with the critical ratio which throw some light upon the differences which were obtained between the means of the various tests. Observation will show that, in general, the changes that occur

TABLE I*

Boys..... 39 Subjects

Based on Total Scores (Paragraph Comprehension Plus Word Meaning)

	Non-Radio I	Radio I	Radio II	Radio III	Radio IV	Radio V	Non-Radio II
Range	(121–194)	(130–207)	(72–218)	(74–219)	(83–202)	(86–194)	(99–206)
Mean	158.27	165.19	166.22	151.99	154.42	151.35	155.58
SD	17.70	20.3	24.4	28.35	27.95	24.85	25.70
Mean diff	+6.92	+1.03	−14.23	+2.43	−3.07	+4.23	
SD diff	9.57	17.3	24.43	16.88	18.01	19.73	
SE diff	1.55	2.8	3.95	2.74	2.92	3.19	
Critical Ratio	4.46	.37	3.78	.89	1.05	1.32	
Up 4 or more points				24	18	7	18	13	19
Down 4 or more points				6	11	28	10	14	15
Changed less than 4 points				9	10	4	11	12	5

* Comparisons are made thus; Radio I over Non-Radio I, Radio II over Radio I, Radio III over Radio II, Radio IV over Radio III, Radio V over Radio IV, Non-Radio II over Radio V.

TABLE II

Girls..... 46 Subjects

Based on Total Scores (Paragraph Comprehension Plus Word Meaning)

	Non-Radio I	Radio I	Radio II	Radio III	Radio IV	Radio V	Non-Radio II
Range	(112–196)	(125–211)	(104–211)	(64–191)	(119–205)	(114–205)	(72–201)
Mean	158.59	160.98	166.85	150.87	155.33	156.30	149.40
SD	18.35	20.45	23.75	24.05	18.80	21.75	27.30
Mean diff	+2.39	+5.87	−15.98	−4.46	+.97	−6.90	
SD diff	14.79	16.25	16.84	17.43	12.67	20.66	
SE diff	2.20	2.42	2.51	2.59	1.89	3.08	
Critical Ratio	1.09	2.43	6.37	1.72	.52	2.24	
Up 4 or more points				20	28	5	24	22	18
Down 4 or more points				18	10	37	13	16	20
Changed less than 4 points				8	8	4	9	8	8

between the seven means of each table are roughly comparable to the changes occurring between the seven medians shown in Figures I and II.

There is, for both boys and girls, a rise between the mean of NR. 1 and that of R. 1. The rise is sharper in Table I—Boys, than in Table II—Girls, but both rise. In both there is also a continuation of this rise to a high at R. 2. Furthermore, as was the case with most of the median lines, there is a sharp drop between the means of R. 2 and R. 3. In both tables, the mean of R. 3 is practically as low as any other recorded, the other low means being R. 5 for boys and NR. 2 for girls. The mean of Table I—Boys shows a substantial rise from R. 5 to NR. 2, while that of Table II—Girls, shows an even more substantial drop at the same point.

The obtained critical ratios of Table I show that two of the mean differences are unquestionably reliable. One is so slight as to be unreliable; that is the difference between the means of R. 1 and R. 2. The critical ratio here is .37 which gives only about sixty-five chances out of one hundred that the true difference is above zero and in the obtained direction. This is only slightly above the indications of pure chance.

The last three critical ratios of the same table are considered as of suggestive reliability. They range from eighty-one chances

TABLE III

Boys.....39 Subjects

Interrelationships of Various Scores of Individuals

Correlation between Scores on Non-Radio I and Radio I	r = .88
Correlation between Scores on Radio V and Non-Radio II.....	r = .75
Number of Subjects Scoring above Q3 on Entire Series	5
Number of Subjects Scoring below Q1 on Entire Series	4

in one hundred to ninety-one chances in one hundred that the true differences are above zero and in the obtained direction. While these critical ratios indicate a tendency more significant than that of pure chance, they still fall far short of the critical ratio of 3.0 insisted upon by Garrett (*Statistics in Psychology and Education*, 1938 Revision, p. 231) or even the one of 1.73 accepted by Peters (*Statistical Procedures and Their Mathematical Bases*, 1940 Revision, pp. 476–477). The reader must decide for him-

self how much he shall qualify the descriptive statement of 'suggestive reliability.'

At the bottoms of both Tables I and II are presented data indicating the degree to which recorded changes were influenced by few or by many of the subjects. These results are arranged to show how many pupils increased four or more points, how many decreased their scores four or more points, and how many changed their scores less than four points. Four points was chosen since the PE of a score is given by the Stanford Test authors as about two points for Grade VIII.

The increase from NR. 1 to R. 1 was shared in by twenty-four out of the thirty-nine boys, while only six dropped four or more points. This indicates that the change in both mean and median was caused by many small increases rather than by a few large ones. Similarly, the abrupt drop between R. 2 and R. 3 was shared in by twenty-eight of the thirty-nine subjects. This likewise would indicate a small drop by many pupils rather than large decreases by a few.

In Table II—Girls, the difference between means increases slightly but is scarcely reliable and the median drops slightly.

As in observable in all the data, there is a tendency to 'level off' after R. 3. The figures indicating both change and stability for the remainder of the tests tend to be reasonably constant.

TABLE IV

Girls.....46 Subjects

Interrelationships of Various Scores of Individuals

Correlation between Scores on Non-Radio I and Radio I.	r = .70
Correlation between Scores on Radio V and Non-Radio II.....	r = .71
Number of Subjects Scoring above Q3 on Entire Series.	7
Number of Subject Scoring below Q1 on Entire Series..	5

Tables III and IV are designed to show some interrelationships of scores made by individuals in the various test periods. One way of studying this was to determine the correlations for both boys and girls between the first non-radio test and the first radio test, in one instance, and between the fifth radio test and second non-radio test, in another. This was done to see to what extent pupils tended to hold their relative positions on both type tests. Tables III and IV show that there is a

substantial difference between the coefficients of correlation for scores on NR. 1 and R. 1 of boys and girls. The r for boys is .88 and that for girls, for the same tests, is .70. However, between R. 5 and NR. 2, the r's show no significant difference. The r for boys is .75 and that for girls is .71. This would indicate that the girls tended to change relative positions between NR. 1 and R. 1 more than did the boys. In addition to the correlations, the data were studied to ascertain if there was a possible tendency of certain pupils to maintain levels either above Q. 3 or below Q. 1 in all test periods. It was found that nine boys out of a total of thirty-nine maintained this relatively stable position, while twelve girls out of forty-six did so.

The experimenter was interested in determining which individuals contributed most to that rise and if any particular reading ability level had more to do with bringing it about; in other words, did the better readers or the poorer ones contribute most? A twenty-point increase was selected arbitrarily as the minimum to be considered. These data (not presented here) indicate that the sharp increases come from widely separated reading ability levels as measured by the initial test. In the boys' group, this is especially true; the range of initial scores, on the part of those making the substantial increases from NR. 1 to R. 2, is from 121 to 181. The mean of NR. 1, Boys, was 158.27. Thus we found that four of the pupils gaining twenty or more points were below the mean and seven above. This might indicate that among the boys the better readers contributed more to the rise from NR. 1 to R. 2.

It was found that a smaller proportion of girls (eight out of forty-six) gained twenty or more points between NR. 1 and R. 2 than of the boys (eleven out of thirty-nine). Furthermore, there is a tighter range in the girls making this increase (148–180) than in the boys making similar increases (121–181). Seven of the eleven boys were above 158 in initial score, while five of the eight girls scored 158 and above. Of the remaining four boys, three were below the Q. 1 (143) of their group. The remaining three girls scored 156, 152, and 148, respectively. (Q. 1 of girls equals 149.)

In the case of both boys and girls, only two students who gained twenty or more points were in the upper quarters of their groups. Thus the majority of cases making this large gain were

in the middle two-thirds of their group, with a tendency for more of them to be above the median rather than below. In the case of the girls, no extremely poor readers made this gain of twenty points, while three of the boys were definitely in the lowest quarter on the initial test. One of these was the lowest boy on the initial test.

CONCLUSION

Perhaps the most apparent conclusion to be drawn from this experiment is that there is not a great deal of positive evidence. Trends might be shown or tendencies indicated, but the study might indicate some fields of experimentation for others. Much could be done concerning the differences in effects on study efficiency of radio music or radio speech. Phonographs and recordings could be employed for such research. It might also be profitable for someone to attempt to determine the relationship between degree of intelligence and effect of radio listening upon study efficiency. Perhaps, it might be worth while to conduct experiments dealing with ability to concentrate at various grade levels in the presence of radio. The author feels that the chief value of this study is to stimulate further experimentation on the part of other workers.

Since the object of the study, as stated earlier, was to determine the effect of radio-listening on pupils' ability to concentrate on study-type materials, the answer to this aim would be the crux of any conclusions drawn from the experimental data.

In spite of slight variations which occur from test to test, it is the writer's judgment that, by and large, this study presents little substantial evidence indicating that radio listening of the type used either aids or impedes ability to study. As a whole, both boys and girls did just about as well with the radio playing as they did without it.

The boys made a significant increase in score form NR. 1 to R. 1. This might mean that radio listening aided study ability. The fact that this rise is carried on to R. 2 and that the girls showed an increase at R. 2 would add emphasis to this conclusion; or, as previously mentioned, practice effect might have entered. But the great drop from R. 2 to R. 3 does much to cancel these gains. A study of both means and graph lines indicated that beyond R. 3, gains and losses tend to balance each other.

The fact that the boys scored gains from NR. 1 to R. 1 while the girls showed practically no change at the same point might indicate that switching from a quiet classroom situation to one with a radio playing stimulated boys more than girls. On the other hand, it could indicate that the conducting of an experiment caught the imagination of the boys more than that of the girls.

The abrupt drops from the high of R. 2, shown in means and in all graph lines, coupled with the fact that in almost no case do either means or lines again approach these high points, would indicate that beyond R. 2 the novelty of the experimental situation had worn off and that the radio was having little or no effect on study efficiency.

In summary, the conclusions might be stated thus:

1) There is little or no significant difference apparent in efficiency of reading-type study directly ascribable to presence or absence of radio.

2) What slight differences appear seem to indicate that the boys were affected more quickly by the experimental situation than were the girls.

3) The results of this study indicate that considerably more research should be done in this field.

THE SPEED OF OBJECT RECOGNITION AND OF WORD RECOGNITION IN GROUPS OF PASSING AND FAILING PUPILS

THOMAS H. EAMES

School of Education, Boston University

Since word recognition is a factor in reading ability and the speed of word recognition is one of the factors in the speed of reading, it was thought that a study of recognition speed in passing and failing groups of pupils might contribute to the understanding of the causes of failure in reading. Accordingly three hundred twenty-nine cases were studied, two hundred fifty-four of whom were failing in reading or in subjects requiring reading, and seventy-five of whom were passing in all subjects.

The technique followed was that reported in the *Journal of Educational Research* in 1937[1] and recapitulated briefly in the *American Journal of Ophthalmology* in 1938.[2] This consisted of the projection in measured flashes at book distance of both pictures of objects and familiar words known to be in the subject's sight vocabulary. The author has used this technique since before 1935. In recent years, especially since 1942, commercially-made tachistoscopes involving the same principle have appeared in increasing numbers. Results in short exposure studies are not closely comparable unless made with the same model instrument with the same illumination, at the same screen distance and with stimulus pictures and words which are similar, especially as regards style of type and size. The use of different styles of tachistoscopes (which in most cases seem to the author to reflect an engineering rather than a pedagogical background on the part of the designer) appears likely to yield not very comparable mathmatical readings, and so it is urged that general trends be considered rather than actual measurements of time unless the apparatus and test situations are known to be closely similar.

---

[1] Eames, T. H. "A Study of the Speed of Word Recognition," *Jour. Educ. Res.*, 31: (Nov.) 1937.

[2] Eames, T. H. "The Speed of Picture Recognition and the Speed of Word Recognition in Cases of Reading Difficulty." *American Jour. Ophthal.*, 21: (Dec.) 1938.

Measurements were made of the speed of object (picture) recognition and the speed of word recognition in order to make possible a comparison of a literate with a non-literate measure of recognition speed. The cases were divided into passing and failing groups and were distributed under these headings. Central tendencies were calculated for each group and the two median speeds were compared within each group as well as from group to group. The speed ceiling of the tachistoscope used was .0033 second and several medians fell in this upper interval indicating that, in such instances, at least half of the cases in the group had speeds of this amount or faster.

The passing group as a whole presented medians of .0033 plus second in both the speed of object recognition and the speed of word recognition, while the failing group presented a median speed of object recognition of .0033 plus second and a median speed of word recognition of .0126 second.[3] The two groups were then redistributed according to age instead of grade. This was done because most members of the failing group had been retarded at least one year, and hence a consideration based on grade would be misleading in the evaluation of results from the physiologic as well as the psychologic viewpoint. The passing group presented median speeds of object recognition at the upper level throughout the entire age range from six through thirteen. The same group displayed similar median speeds of word recognition through the ninth year, after which there was a sharp regression in the tenth year, with recovery at the eleventh year which was maintained throughout the rest of the range.

On the other hand, the failing group presented median speeds of object recognition within the .0033 second plus interval at all ages, but exhibited a considerably slower initial speed of word recognition, which remained stationary at the sixth- and seventh-year levels, rose steadily at eighth and ninth, then, like the passing group, fell off at tenth but, unlike the latter, made only slight recovery at the eleventh- and twelfth-year levels. The median speed of word recognition did not regain the level of the ninth year until the thirteenth year, at which point a sharper rise carried

[3] This is closely comparable with the results of a study of four hundred twelve failures in which group the median speed of object recognition was .0033 plus second and the median speed of word recognition was .0100 second.

the median into the maximum speed interval of .0033 plus second.

The fact that the median speed of object recognition occurred in the maximum speed interval in both groups, while the median speed of word recognition exhibited various degrees of retardation, especially in the failing group, emphasizes educational factors in the etiology. It is particularly noteworthy that both passing and failing groups exhibited a regression in the median speed of word recognition at the tenth-year level and that, although the passing pupils recovered their median level by the next year, the failing group presented a plateau effect with rather slight yearly increment until a sharper increase at the thirteenth year carried the median to the upper level.

TABLE 1.—COMPARISON OF GROUP AND AGE MEDIANS

Age	Passing		Failing	
	Object	Word	Object	Word
6	.0033+	.0033+	.0033+	.0400
7	.0033+	.0033+	.0033+	.0400
8	.0033+	.0033+	.0033+	.0270
9	.0033+	.0033+	.0033+	.0050
10	.0033+	.0014+	.0033+	.0113
11	.0033+	.0033+	.0033+	.0133
12	.0033+	.0033+	.0033+	.0109
13	.0033+	.0033+	.0033+	.0033+
all	.0033+	.0033+	.0033+	.0126
Number cases	75		254	Total 329.

This suggests many possibilities, but the author believes a probable explanation to be as follows: During the sixth-, seventh-, and eighth-year levels oral reading instruction is emphasized, probably ranging in the majority of schools from seventy per cent to ninty per cent of the total time devoted to reading. During the period the development of sight vocabulary and the general mechanics of reading are stressed. As the child goes from grade to grade the amount of time devoted to the mechanics

and oral reading is diminished until at about the ninth year of age the emphasis shifts to practical silent reading for meaning, such as the reading of problems in arithmetic and the use of texts and assignments in the social studies such as geography. The child having been accustomed to the teacher's close supervision and help in oral reading, finds himself much more on his own. His silent reading errors often go unrecognized and impair meaning. This necessitates a readjustment in most children's attitude toward reading, results in some degree of confusion and insecurity in some cases, and accounts for the drop in recognition speed for words in both passing and failing groups. The failing group, being the more insecure, readjusts more slowly than the passing group and takes longer to adapt itself. This is probably not the only explanation, but it seems well worth considering.

If the above deductions are correct, the figures tend to emphasize the more recent pedagogical tendency to encourage teachers at all grade levels to teach reading, not perhaps as a specific subject, but as a part of general instruction. They also indicate the advisability of urging third- and fourth-grade teachers to take more careful steps to cushion the change from instruction in the mechanics of reading to the practical use of the subject in other areas. Furthermore, they should encourage teachers to individualize their reading instruction as much as possible and to move farther and farther away from the less desirable features of the grade system with its attendant lockstep in both curriculum and instruction.

BOOK REVIEWS

EMMETT A. BETTS. *Foundations of Reading Instruction.* New York: American Book Co., 1946, pp. 757.

The central theme of this text, designed for use in teachers colleges and schools of education, is concerned with how to identify individual needs and how to provide for them in the classroom situation. Major emphases are differential guidance, general language development, reading readiness, semantic or meaning basis of language, social basis of language, and systematic sequences. General sections of the book deal with the reading situation, the reading problem, the nature of readiness, developing readiness, and reading instruction.

It is obvious that a review cannot cover systematically the material in a book of this size. The discussion will be limited, therefore, to a sampling of what the reviewer considers to be some of the strong and weak points. Among the more noteworthy contributions are the following: (1) The emphasis upon differentiated guidance is necessary if we are to prevent the majority of our present reading ills. (2) Readiness for reading activities is necessary at all levels in a developmental reading program. (3) In order to emphasize that reading is a process, it is desirable to stress 'reading-to-learn' rather than 'learning-to-read.' This emphasizes that reading is a thinking process and that basic consideration should be given to meaning. (4) The successful reader is the versatile reader, i.e., he can change his pace to fit the purpose and content of the reading material. (5) Purpose in reading is adequately stressed. The formation of reading purposes should be largely a teacher and class coöperative affair. (6) Irrespective of whether physical disability is related to reading proficiency or not, disability should be corrected as far as that is possible. (7) The important rôle of the teacher's judgment in determining readiness for initial reading instruction is given proper emphasis. (8) Important cautions are outlined concerning the use of workbooks and basal readers. (9) The discussion of aids to word recognition and meaning is one of the best the reviewer has seen. (10) The emphasis upon four basic types of information needed for clinical appraisal is desirable: (a) basal or independent reading level; (b) instructional level, or highest reading level at which systematic instruction can be initiated;

(c) the level of frustration in reading, or the level at which the individual is baffled or thwarted by the language; and (d) probable capacity level, or highest reading level at which the individual can comprehend material read to him. (11) The many well-chosen photographs of classroom activities are informative. (12) The references are complete.

Following are listed certain criticisms of the book: (1) On page 75 it is stated that fixation pauses require about ninety-six per cent of the total reading time. Results show that it is about ninety-three per cent. (2) It is stated (p. 333) that a very good reader will make three or four fixation pauses per line and that the good reader can read four hundred to twelve hundred words each minute. The reviewer doubts that any person can read twelve hundred or even nearly twelve hundred words per minute. With ten words per line and three to four fixations per line, this would mean that the average pause duration would have to be only about .170 to .125 second. Experiments show that for good adult readers, fixation pause durations while reading easy prose average about .220 to .250 second. The latter would mean 1.45 to 2.00 minutes for reading the twelve hundred words, rather than one minute. (3) Certain material is included that adds little or nothing to the main trend of the discussion, i.e., Chapters II, III, and IV. Too much space is devoted to argument for curriculum and educational reform. (4) Although the author is apt in his choice of quotations, he lessens the effectiveness of his discussion by too frequent use of lengthy citations. In one chapter, for instance, more space is taken up by quoted material than by the author's discussion. (5) The rather constant tendency to present material in outline form is disconcerting to the reader. The excessive detail involved makes it difficult to follow the general trend of the discussion. (6) Detailed statements of methodology, such as given in describing the measurement of visual and auditory deficiency, are not essential since this material is always available in manuals. (7) Quite generally, the organization of the author's discussions is not well done. There are to many ideas and views thrown together in a somewhat uncoördinated manner. Also there is considerable repetition.

How then should the book be evaluated? For the most part, viewpoints of the author are up to date and basically sound. The book is 'packed' with facts, sound views and practical sug-

gestions. But, because of its organization, it is a very hard book to read. And the book is long—seven hundred fifty-seven large, double-column pages. Obviously the book has limitations for use as a text in courses on reading. However, the reviewer would suggest that every elementary-school teacher can profit by having the book available as a reference work. Much of the worth-while information in the field will be found somewhere in the book. Also the bibliographies of journal articles, books, tests, etc. will be found very useful.

MILES A. TINKER
University of Minnesota

ABRAM KARDINER, RALPH LINTON, CORA DUBOIS AND JAMES WEST—Collaborators. *The Psychological Frontiers of Society.* New York: Columbia University Press, 1945, pp. 475.

In his latest book Kardiner continues to work toward the goal he set for himself in *The Individual and His Society* (1939), that of developing a 'science of society' and perfecting a technique for "tracking down the reciprocal relations between culture and personality." Toward this end, he has collaborated with such social scientists as Linton, West and DuBois, his own contribution to this synthesis being his application of a psychoanalytical conceptual scheme to ethnographic studies they had made.

But this is not orthodox Freudian psychoanalysis. For, substituted for Freud's 'instincts' are 'action systems' with recognizable cognative, coördinative and executive functions. Nevertheless, psychoanalysis is preferred to other schools of psychology for the study of ethnographic data because it is, at once, dynamic, genetic and holistic. Present-day behaviorism is rejected for its atomistic and mechanistic approach to the organism and because it is incapable of dealing with the projective phenomena which figure so prominently in human adaptation. Gestalt psychology is seen as having little value to social science at the present time because it has a yet only one dynamic principle, that of dynamic inter-action. Kardiner views Lewin's topology as an outgrowth of Gestalt psychology and states that its success is not established and that its problems are largely those defined by psychoanalysis.

As in his earlier work, Kardiner studies and interprets psychoanalytically, ethnographic data compiled about several cultures

to derive the 'basic personality structure' of each. The basic personality type for any culture is defined as that "personality configuration which is shared by the bulk of the members of that culture as a result of the early experiences which they had in common."(viii) A list of key situations which influence personality formation is presented. But despite the comprehensive nature of the list, it is quite apparent that experiences undergone during the formative years are considered the most important factors molding basic personality.

Fundamental components of the 'basic personality structure' are 'projective systems' and 'taught reality systems.' The latter consist of learning imparted by direct tuition about the outer world, how to deal with it, man's relationship to it and the conventionalized attitudes which govern the relationships of men to each other. The former, which generally take the form of religion, ritual, folklore, taboo systems, etc., develop as a means of relieving constellations of needs and tensions created by the primary institutions (feeding, weaning, maternal care, basic disciplines, family organization, etc.) which the individual cannot control and which create problems of adaptation. Further, the 'projective systems' exercise a polarizing influence over all 'taught reality systems.' This means simply that before innovations will be accepted in a culture, they must be plausible and tenable in terms of the experiences the individual had during his formative years and they must be consistent with the beliefs, attitudes and philosophy he had developed during that period.

The cultures which are presented in *The Psychological Frontiers*, selected because their personality configurations differ so markedly, are Comanche, Alor and Plainville, U.S.A., an American rural community. In the case of the people of Alor and those of Plainville, Kardiner, in addition to the usual ethnographic data, had access to biographical sketches and dream material. He also had thirty-eight Rorschach protocols, word association tests, Porteus test results and children's drawings which DuBois had collected during her study of the Alorese.

Kardiner dismisses the word association tests and the children's drawings as not adding any information not available through the application of his own technique. But he recommends continued use of the Porteus Intelligence Test or something similar. The biographical material, he found, corroborated,

most respects, his own initial conjecture regarding the 'basic personality structure' of the Alorese. But he discovered that his own technique, based on a study of institutions rather than individuals, was not refined or sensitive enough to pick up the difference in character formation between the male and female members of the culture. From this he concludes that "personality studies are indispensable for the coördination of the institutional configuration."(248) The Rorschach protocols (blind analyses of which were made by Emil Oberholzer) provided startling confirmation of Kardiner's own technique. However, he sees in the Rorschach more than just a validation of his own approach. He considers it a valuable and necessary adjunct because it makes possible the diagnosis of "the constellations existing at the periphery of the personality, some of which might be passed up in the projective analysis."(249)

Although Kardiner would probably protest to the contrary, one gets the impression that he did not limit himself to a psychoanalytical approach, but rather that his approach is an eclectic one. For throughout his work are references to the conditioning process and his very use of behavioristic concepts, which he criticizes, seems proof of the fact that psychoanalytical concepts do not suffice to satisfactorily explain all aspects of behavior and experience. But despite his refusal to acknowledge a debt to systems of psychology other than Freud's, there is no doubt but that Kardiner has provided social science with a tool that lays low, once and for all time, the concepts of 'instinct' and racism. This tool has long been overdue.

While Kardiner's technique for deriving 'basic personality structure' is essentially a valid one and his preoccupation with the early experiences of the formative years is sound, it would seem that he makes an error in assuming that the child-rearing techniques within a single society are similar. This may be true of simple pre-literate cultures. But it certainly is not true of our own where class-stratification is highly developed, where there are caste-like and ethnic subgroups, and, where the life of the urban dweller differs so markedly from that of those living in rural communities. Sociological studies have provided ample evidence to the fact that there is wide difference in the child-rearing practices by the several different American classes and that these practises are also related to religious differences.

In his last chapter, "Basic Personality and History," Kardiner makes an attempt to analyze the history of Western civilization in accordance with the concepts of 'basic personality structure.' He concludes that the primary institutions have not changed much in configuration though they have changed in content and that the projective systems in Western society, notably religion, have been divested of their power to influence social stability by the reality systems. This he considers of paramount importance for religion has served as a stabilizing influence by externalizing the basic disciplines which prevailed in the community. Such use of religion is considered but a poor substitute for the reality of sound and wholesome interpersonal relations and Kardiner implies that unless society can foster such factors as will create a type of personality conducive to higher degrees of coöperation, the elimination of mistrust, suspicion and mutual hostility, its stability and our present way of life will be seriously threatened.

Notwithstanding the usual shortcomings typical of most psychoanalytical literature, such as roving and digressive style, this book should be read widely for it is in the vanguard of what promises to be the first important step in the integration of the social sciences concerned with human behavior.

STANLEY LIPKIN

The University of Chicago